

---

# **Wait... what do you mean by WikiCite?**

James Hare • WikiCite Monthly Meeting • 2023-01-31

---

## Good question. Let's start from the beginning.

*First version of "[WikiCite 2016](#)" on Meta, DarTar, 12 February 2016:*

"WikiCite 2016 is **an event** focused on technology to improve the coverage, quality, standards-compliance and machine-readability of **citations and source metadata in Wikipedia**, Wikidata and other Wikimedia projects."

*First (non-redirect) version of "[WikiCite](#)" on Meta, DarTar, 5 September 2016:*

"WikiCite is a proposal to build **a bibliographic database in Wikidata** to serve all Wikimedia projects."

---

## Event? Database? Both?

There were more issues we needed to address as well:

- Citation *events* (X article cites Y work) vs. source metadata (details about Y work)
- Storage in Wikidata vs. another Wikibase vs. another thing entirely
- WikiCite as a “tenant” of Wikidata vs. WikiCite as a benefit to Wikidata
- Use cases for a reference database not directly related to the original goal of tracking citations on Wikipedia e.g. Zika corpus
- And then all the work that went into building the database itself, within Wikidata, including bots and scripts

---

## Major projects from WikiCite

- Cite Q – cite a source in a Wikipedia article via that source's Wikidata ID
- Inventaire, Scholia – frontends for data stored in Wikidata
- Initiative for Open Citations – freed up journal article citation data
  - Subsequent development of rich citation graph in Wikidata
- Shared Citations – proposed bottom-up database of citation events
- Several projects to build the WikiCite corpus on Wikidata

---

## A partial summary of results to-date

- 42 million Wikidata **items about publications**
  - Many of which correspond to publications cited in Wikipedia
  - ...and many that don't!
- 287 million **citation relationships** described within Wikidata
- 42,000 uses of **Cite Q** on English Wikipedia
- Three standalone WikiCite conferences and 19 other WikiCite-events
- 12 **grant-funded projects** on top of several community projects
- This is not a complete list

---

## But what about the original goals?

“WikiCite is a proposal to build a ***bibliographic database in Wikidata*** to serve all Wikimedia projects.”

- kind of? The bibliographic database is definitely there, but its scope is varied

- growth is constrained by Wikidata’s limitations

“WikiCite 2016 is ***an event*** focused on technology to improve the coverage, quality, standards-compliance and machine-readability of ***citations and source metadata in Wikipedia***, Wikidata and other Wikimedia projects.”

- the events have been successful in building grassroots interest and were expected to grow in number and variety before a once-in-a-century pandemic got in the way

- that said, there has been more emphasis around *source metadata* than ***citations***.

---

# Challenges

- We have a strong baseline in the current Wikidata corpus, *but*:
- Further growth is limited by the social and technical limitations of Wikidata
  - Query service timeouts
  - Ability to search and organize data at massive scale
- Funding for WikiCite ended in 2021; Shared Citations not funded
- Our in-person events, where large bursts in progress happened, largely stopped
- And all these issues have combined to **stall overall momentum**

---

## Some solutions

- **Internet Archive is stepping up in 2023**
  - Organizing monthly meetings
  - Wikipedia Article Reference Inventory and Wikipedia Citations Database
- Separately, I have been working on some projects
  - Orb Open Graph, WikiCite-friendly Wikidata Query Service alternative <<https://orb.how>>
  - I want to bring back Librarybase in 2023. This will be a gradual process.
- Triumphant return of in-person events
  - WikiCite participation in Athens hackathon, Singapore Wikimania
  - Potential WikidataCon satellite event at Internet Archive in San Francisco
  - WikiCite 2024?



---

## Let's simplify the terminology

- If **Wikidata** is the Wikibase (currently) used for source metadata
  - and **Librarybase** a potential supplement
- and **Shared Citations** was the proposed database for citation events
  - with **Wikipedia Citations Database** as a prototype

Then I propose we use **WikiCite** to refer to the events and community that made this all happen.

<https://meta.wikimedia.org/wiki/WikiCite> (you can edit this)

---

# WikiCite and the Internet Archive

James Hare • WikiCite Monthly Meeting • 2023-01-31

---

# What is the Internet Archive?

- Nonprofit organization founded in San Francisco in 1996
- Archiving the world wide web for over 25 years
- Collections of tens of millions of digitized books, periodicals, multimedia, emulated software, and more
- Universal access to all knowledge

<https://commons.wikimedia.org/wiki/File:2018-1130-internet-archive-visit-by-wikimedians-05.jpg>



---

# Wayback Machine and Turn All References Blue

<https://web.archive.org>

Crawling the web and taking snapshots 24/7

All external links posted to all Wikimedia wikis have been submitted for archiving since ~2013.

Primary source for **InternetArchiveBot**, a bot that has fixed over 15 million broken links on Wikimedia projects

**Turn All References Blue** builds on the link-fixing work of InternetArchiveBot:

- Improve the references that already exist on the wikis
- Make it easier for Wikimedia contributors to cite Internet Archive holdings or incorporate them into the wikis
- Build linked open databases around works that are, or could be, cited on-wiki

---

## Internet Archive's role in WikiCite

- WikiCite is an open garden which we have started to tend to
- Wikipedia needs sources and Internet Archive has sources
- We need better infrastructure for source discovery
- Our team has experience with Wikipedia and Wikidata
- We would like to open up our space to events

---

# Our projects supporting WikiCite

- Monthly meetings and participation in events
- Enhancing references with links to books and periodicals
- **Wikipedia Citations Database**  
<https://wikicitations.wikibase.cloud>
  - Prototype Wikibase of citation events (X article cites Y resource)
- **Wikipedia Article Reference Inventory**
  - Snapshot-in-time view of citations on a single article

```
"wikitext": "<ref>{{cite web |last=Dangerfield |first=Whitney |title=The Mystery of Easter Island |url=https://www.smithsonianmag.com/travel/the-mystery-of-easter-island-151285298/ |website=[Smithsonian (magazine)|Smithsonian Magazine]] |access-date=December 10, 2020 |date=March 31, 2007}}</ref>",
"is_citation_reference": true,
"is_general_reference": false,
"has_archive_details_url": false,
"has_google_books_url_or_template": false,
"has_web_archive_org_url": false,
"url_found": true,
"doi": "",
"isbn": "",
"urls": [
  {
    "checked": true,
    "request_error": false,
    "first_level_domain": "smithsonianmag.com",
    "first_level_domain_done": true,
    "fld_is_ip": false,
    "malformed_url": false,
    "dns_record_found": true,
    "scheme_missing": false,
    "added_http_scheme_worked": false,
    "dns_no_answer": false,
    "no_netloc": false,
    "request_url_error": false,
    "unrecognized_scheme": false,
    "unrecognized_tld": false,
    "parsing_done": true,
    "dns_error": false,
    "status_code": 200,
    "url": "https://www.smithsonianmag.com/travel/the-mystery-of-easter-island-151285298/",
    "fixed_url": "",
    "scheme": "https",
    "netloc": "www.smithsonianmag.com",
    "tld": "com"
  }
]
```